

Learning Optimal Social Dependency for Recommendation

Yong Liu, Peilin Zhao, Xin Liu, Min Wu, Xiao-Li Li

Institute for Infocomm Research, A*STAR, Singapore

{liuyo, zhaop, liu-x, wumin, xlli}@i2r.a-star.edu.sg

Abstract

Social recommender systems exploit users' social relationships to improve the recommendation accuracy. Intuitively, a user tends to trust different subsets of her social friends, regarding with different scenarios. Therefore, the main challenge of social recommendation is to exploit the optimal social dependency between users for a specific recommendation task. In this paper, we propose a novel recommendation method, named probabilistic relational matrix factorization (PRMF), which aims to learn the optimal social dependency between users to improve the recommendation accuracy, with or without users' social relationships. Specifically, in PRMF, the latent features of users are assumed to follow a matrix variate normal (MVN) distribution. The positive and negative dependency between users are modeled by the row precision matrix of the MVN distribution. Moreover, we have also proposed an efficient alternating algorithm to solve the optimization problem of PRMF. The experimental results on real datasets demonstrate that the proposed PRMF method outperforms state-of-the-art social recommendation approaches, in terms of root mean square error (RMSE) and mean absolute error (MAE).

1 Introduction

Recommender systems have been widely used in our lives to help us discover useful information from a large amount of data. For example, on the popular e-commerce websites such as Amazon, the recommender systems predict users' preferences on products based on their past purchasing behaviors and then recommend a user a list of interesting products she may prefer [Linden *et al.*, 2003]. For a recommender system, the most critical factor is the prediction accuracy of users' preferences. In practice, the most successful prediction methods are collaborative filtering based approaches, especially the matrix factorization models [Su and Khoshgoftaar, 2009].

The recent rapid developments of online social networking services (e.g., Facebook and Twitter) motivate the emergences of social recommender systems that exploit users' online social friendships for recommenda-

tion [Yang *et al.*, 2014]. The social recommender systems usually assume that a user has similar interests with her social friends. Although the recommendation accuracy can usually be improved, there do not exist strong connections between the online friendship and the similarity of users' interests [Ma, 2014]. Because there are different categories of social networking friends in online social networks, e.g., school friends, work-related friends, friends sharing same interest/activities, and neighborly friends [Zhang *et al.*, 2013]. The tastes of a user's online friends usually vary significantly. In different scenarios, a user tends to trust the recommendations from different subsets of her online social friends. Hence, the key to success for a social recommender system is to exploit the most appropriate social dependency between users for recommendation.

Moreover, existing social recommender systems are usually developed based on users' explicit social relationships, e.g., trust relationships [Jamali and Ester, 2009; Jamali and Ester, 2010] and online social friendships [Ma *et al.*, 2011; Yang *et al.*, 2012; Tang *et al.*, 2013]. They ignore the underlying implicit social relationships between users that have most similar or dissimilar rating behaviors. This implicit social relationships have been found to be beneficial for improving the recommendation accuracy [Ma, 2013]. In addition, users' explicit social relationships may be unavailable in many application scenarios. This also limits the application of traditional social recommendation approaches.

In previous studies, users' social dependency adopted for recommendation are predefined based on users' explicit or implicit social relationships [Ma, 2013; Yang *et al.*, 2014]. Differing from previous work, this paper proposes a novel social recommendation method, namely probabilistic relational matrix factorization (PRMF), which aims to learn the optimal social dependency between users to improve the recommendation accuracy. The proposed method can be applied to the recommendation scenarios with or without users' explicit social relationships. In PRMF, the user latent features are assumed to follow a matrix variate normal (MVN) distribution. The positive and negative social dependency between users are modeled by the row precision matrix of the MVN distribution. This is motivated by the success of using the MVN distribution to model the task relationships for multi-task learning [Zhang and Yeung, 2010]. To solve

the optimization problem of PRMF, we propose a novel alternating algorithm based on the stochastic gradient descent (SGD) [Koren *et al.*, 2009] and alternating direction method of multipliers (ADMM) [Boyd *et al.*, 2011] methods. Moreover, we extensively evaluated the performances of PRMF on four public datasets. Empirical experiments showed that PRMF outperformed the state-of-the-art social recommendation methods, in terms of root-mean-square error (RMSE) and mean absolute error (MAE).

2 Related Work and Background

In this section, we first introduce some background about probabilistic matrix factorization (PMF) [Mnih and Salakhutdinov, 2007], one of the most popular matrix factorization models. Then, we review the state-of-the-art social recommendation methods.

2.1 Probabilistic Matrix Factorization

For the recommendation problem with m users $\{u_i\}_{i=1}^m$ and n items $\{v_j\}_{j=1}^n$, the matrix factorization models map both users and items into a shared latent space with a low dimensionality $d \ll \min(m, n)$. For each user u_i , her latent features are represented by a latent vector $U_i \in \mathbb{R}^{1 \times d}$. Similarly, the latent features of the item v_j are described by a latent vector $V_j \in \mathbb{R}^{1 \times d}$. In the PMF model, users' ratings on items are assumed to follow a Gaussian distribution as follows:

$$p(R|U, V, \sigma^2) = \prod_{i=1}^m \prod_{j=1}^n [\mathcal{N}(R_{ij}|U_i V_j^\top, \sigma^2)]^{W_{ij}}, \quad (1)$$

where $R \in \mathbb{R}^{m \times n}$ is the matrix denoting users' ratings on items; $U \in \mathbb{R}^{m \times d}$ and $V \in \mathbb{R}^{n \times d}$ denote the latent features of all users and items, respectively; σ^2 is the variance of the Gaussian distribution; W_{ij} is an indicator variable. If u_i has rated v_j , $W_{ij} = 1$, otherwise, $W_{ij} = 0$. In addition, we also place zero-mean spherical Gaussian priors on U and V as:

$$p(U|\sigma_u^2) = \prod_{i=1}^m \mathcal{N}(U_i|0, \sigma_u^2 I), \quad p(V|\sigma_v^2) = \prod_{j=1}^n \mathcal{N}(V_j|0, \sigma_v^2 I). \quad (2)$$

where I is the identity matrix, σ_u^2 and σ_v^2 are the variance parameters. Through the Bayesian inference, we have

$$p(U, V|R, \sigma^2, \sigma_u^2, \sigma_v^2) \propto p(R|U, V, \sigma^2) p(U|\sigma_u^2) p(V|\sigma_v^2). \quad (3)$$

The model parameters (*i.e.*, U and V) can be learned via maximizing the log-posterior in Eq. (3), which is equivalent to solving the following problem:

$$\min_{U, V} \frac{1}{2} \|W \odot (R - UV^\top)\|_F^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2, \quad (4)$$

where $W \in \mathbb{R}^{m \times n}$ is the indicator matrix, and W_{ij} is the (i, j) element of W . In Eq. (4), \odot denotes the Hadamard product of two matrices, $\lambda_u = \sigma^2/\sigma_u^2$, $\lambda_v = \sigma^2/\sigma_v^2$, and $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

2.2 Social Recommendation Methods

In recent years, lots of social recommendation approaches have been proposed [Yang *et al.*, 2014]. The Social Regularization (SR) [Ma *et al.*, 2011] is one of the most representative methods. The SR method was implemented by matrix factorization framework. The basic idea was that a user may have similar interests with her social friends, and thus the learned latent features of a user and that of her social friends should be similar. The user similarity can be calculated using the Cosine similarity and Pearson correlation coefficient [Su and Khoshgoftaar, 2009]. To improve the performance of SR, [Yu *et al.*, 2011] proposed an adaptive social similarity function. Moreover, as a user tends to trust different subsets of her social friends regarding with different domains, [Yang *et al.*, 2012] introduced the circle-based recommendation (CircleCon) models, which considered domain-specific trust circles for recommendation. [Tang *et al.*, 2013] exploited both the local and global social context for recommendation. [Ma, 2013] extended the SR model to exploit users' implicit social relationships for recommendation. The implicit social relationships were defined between a user and other users that had most similar or dissimilar rating behaviors with her. In addition, [Li *et al.*, 2015] extended SR to exploit the community structure of users' social networks to improve the recommendation accuracy. [Guo *et al.*, 2015] developed the TrustSVD model, which extended the SVD++ model [Koren, 2008] to incorporate the explicit and implicit influences from rated items and trusted users for recommendation. [Hu *et al.*, 2015] proposed a recommendation framework named MR3, which jointly modeled users' rating behaviors, social relationships, and review comments.

3 The Proposed Recommendation model

This section presents the details of the proposed PRMF model that extends PMF to jointly learn the user preferences and the social dependency between users.

3.1 Probabilistic Relational Matrix Factorization

In Section 2.1, the PMF model assumes the users are independent of each other (see Eq.(2)). Thus, it ignores the social dependency between users. However, in practice, users are usually connected with each other through different types of social relationships, *e.g.*, trust relationships and online social relationships. In this work, to exploit users' social dependency for recommendation, we place the following priors on the user latent features:

$$p(U|\sigma_u^2) \propto \left(\prod_{i=1}^m \mathcal{N}(U_i|0, \sigma_u^2 I) \right) q(U), \quad (5)$$

where the first term of the priors is used to penalize the complexity of the latent features of each user, and the second term $q(U)$ is used to model the dependency between different users. Specifically, we define

$$q(U) = \mathcal{MN}_{m \times d}(0, \Theta^{-1}, I), \quad s.t. \quad \Theta \succ 0, \quad (6)$$

where $\mathcal{MN}_{a \times b}(M, A, B)$ denotes the matrix variate normal (MVN) distribution¹ with mean $M \in \mathbb{R}^{a \times b}$, row covariance $A \in \mathbb{R}^{a \times a}$, and column covariance $B \in \mathbb{R}^{b \times b}$; $\Theta \succ 0$ indicates Θ is a positive definite matrix. In Eq. (6), Θ is the row precision matrix (*i.e.*, the inverse of the row covariance matrix) that models the relationships between different rows of U . In other words, Θ describes the social dependency between different users. Thus, Θ is called the social dependency matrix. In this work, for simplicity, we set the column covariance matrix of the MVN distribution as I , which indicates the user latent features in different dimensions are independent. Moreover, we can also rewrite Eq. (5) as follows:

$$p(U|\Theta, \sigma_u^2) = \mathcal{MN}_{m \times d}(0, (\Theta + \frac{1}{\sigma_u^2}I)^{-1}, I). \quad (7)$$

In practice, a user is usually only correlated with a small fraction of other users. Thus, it is reasonable to assume the social dependency matrix Θ is sparse. To achieve this objective, we introduce a sparsity-inducing prior for Θ as follows:

$$p(\Theta|\gamma) \propto \exp(-\frac{\gamma}{2}\|\Theta\|_1), \quad (8)$$

where γ is a positive constant used to control the sparsity of Θ , and $\|\cdot\|_1$ is the ℓ_1 -norm of a matrix. In addition, the sparsity of Θ can also help improve the computation efficiency of the proposed model (see the discussions in Section 3.2).

Moreover, we also assume users' ratings follow the Gaussian distribution in Eq. (1), and add Gaussian priors on the item latent features as in Eq. (5). Through the Bayesian inference, we have

$$p(U, V, \Theta|R, \sigma^2, \sigma_u^2, \sigma_v^2, \gamma) \propto p(R|U, V, \sigma^2)p(U|\Theta, \sigma_u^2)p(\Theta|\gamma)p(V|\sigma_v^2). \quad (9)$$

Then, the model parameters (*i.e.*, U , V , and Θ) can be obtained by solving the following problem:

$$\min_{U, V, \Theta \succ 0} \frac{1}{2\sigma^2} \|W \odot (R - UV^\top)\|_F^2 + \frac{1}{2\sigma_u^2} \|U\|_F^2 + \frac{1}{2\sigma_v^2} \|V\|_F^2 + \frac{1}{2} [\text{tr}(U^\top \Theta U) - d \log |\Theta + \frac{1}{\sigma_u^2}I| + \gamma \|\Theta\|_1], \quad (10)$$

where $|\cdot|$ denotes the determinant of a matrix. In Eq. (10), the model parameters are learned without prior information about users' social relationships.

Incorporating Prior Social Information

Previous studies have demonstrated that user' explicit social relationships [Yang *et al.*, 2014] and implicit social relationships [Ma, 2013] can help improve the recommendation accuracy. These prior social information (*i.e.*, explicit and implicit social relationships) contains prior knowledge about users' social dependency in a recommendation task.

¹ The density function for a random matrix X following the MVN distribution $\mathcal{MN}_{a \times b}(M, A, B)$ is

$$p(X) = \frac{\exp(-\frac{1}{2}\text{tr}[B^{-1}(X - M)^\top A^{-1}(X - M)])}{(2\pi)^{ab/2} |B|^{a/2} |A|^{b/2}}.$$

Let $p_0(U)$ denote the MVN distribution of the user latent features U derived from the prior information as follows:

$$p_0(U) = \mathcal{MN}_{m \times d}(0, \Sigma, I), \quad (11)$$

where $\Sigma \in \mathbb{R}^{m \times m}$ is the *prior row covariance matrix*. If users' explicit social relationships are available, the elements of Σ can be defined as follows:

$$\Sigma_{ik} = \begin{cases} \text{cov}(R_{i*}, R_{k*}) & \text{if } u_k \in \mathcal{F}(u_i) \text{ or } u_i \in \mathcal{F}(u_k) \text{ or } i=k, \\ 0 & \text{otherwise,} \end{cases}$$

where R_{i*} is the i^{th} row in R ; $\text{cov}(x_1, x_2)$ denotes the covariance between two observations; $\mathcal{F}(u_i)$ denotes the set of u_i 's social friends. Moreover, if users' explicit social relationships are unavailable, we define the elements of Σ as follows:

$$\Sigma_{ik} = \text{cov}(R_{i*}, R_{k*}). \quad (12)$$

To exploit the prior information for recommendation, we aim to minimize the distance between the learned MVN distribution $p(U|\Theta, \sigma_u^2)$ and the MVN distribution $p_0(U)$ derived from prior information. This is achieved by minimizing the differential relative entropy between $p(U|\Theta, \sigma_u^2)$ and $p_0(U)$ as follows:

$$\begin{aligned} & \min_{\Theta \succ 0} \int p_0(U) \log \frac{p_0(U)}{p(U|\Theta, \sigma_u^2)} \\ &= \frac{d}{2} \text{tr}[(\Theta + \frac{1}{\sigma_u^2}I)\Sigma] - \frac{d}{2} \log |(\Theta + \frac{1}{\sigma_u^2}I)\Sigma| - \frac{md}{2} \\ &\propto \frac{1}{2} \text{tr}(\Theta\Sigma) - \frac{1}{2} \log |(\Theta + \frac{1}{\sigma_u^2}I)|. \end{aligned} \quad (13)$$

The differential relative entropy in Eq. (13) can be obtained following the ideas in [Dhillon, 2007].

A Unified Model

Considering both the constraints in Eq. (10) and Eq. (13), we formulate the final objective function of the proposed PRMF model as follows:

$$\begin{aligned} & \min_{U, V, \Theta \succ 0} \frac{1}{2} \|W \odot (R - UV^\top)\|_F^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \\ &+ \frac{\alpha}{2} \{ \text{tr}[\Theta(UU^\top + \beta\Sigma)] - (d + \beta) \log |\Theta + \frac{\lambda_u}{\alpha}I| + \gamma \|\Theta\|_1 \}, \end{aligned} \quad (14)$$

where $\lambda_u = \sigma^2/\sigma_u^2$, $\lambda_v = \sigma^2/\sigma_v^2$, and $\alpha = \sigma^2$; β is the parameter that controls the contribution from prior information.

3.2 Optimization Algorithm

The optimization problem in Eq. (14) can be solved by using the following alternating algorithm.

Optimize U and V : By fixing Θ , the optimization problem in Eq. (14) becomes:

$$\begin{aligned} & \min_{U, V} \frac{1}{2} \|W \odot (R - UV^\top)\|_F^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \\ &+ \frac{\alpha}{2} \text{tr}(U^\top \Theta U). \end{aligned} \quad (15)$$

The optimization problem in Eq. (15) can be solved using the SGD algorithm [Koren *et al.*, 2009]. The updating rules used to learn the latent features are as follows:

$$\begin{aligned} U_i &\leftarrow U_i + \theta(\Delta_{ij}V_j - \lambda_u U_i - \alpha\Theta_{i*}U) \\ V_j &\leftarrow V_j + \theta(\Delta_{ij}U_i - \lambda_v V_j), \end{aligned} \quad (16)$$

where $\Delta_{ij} = R_{ij} - U_i V_j^\top$, Θ_{i*} is the i^{th} row of Θ , and θ is the learning rate. Note the SGD updates are only performed on the observed rating pairs $\mathcal{D} = \{(u_i, v_j, R_{ij}) | W_{ij} > 0\}$.

Optimize Θ : By fixing U and V , the optimization problem with respect to Θ is as follows:

$$\min_{\Theta \succ 0} \text{tr}[\Theta(UU^\top + \beta\Sigma)] - (d + \beta) \log |\Theta + \frac{\lambda_u}{\alpha} I| + \gamma \|\Theta\|_1. \quad (17)$$

The optimization problem in Eq. (17) is convex with respect to Θ . Therefore, the optimal solution $\hat{\Theta}$ to Eq. (17) satisfies:

$$(\hat{\Theta} + \frac{\lambda_u}{\alpha} I)^{-1} - \frac{1}{d + \beta} (UU^\top + \beta\Sigma) = \frac{\gamma}{d + \beta} \hat{G}, \quad (18)$$

where \hat{G} is the sub-gradient, and the elements of \hat{G} are in $[-1, 1]$. Therefore, following the derivation of Dantzig estimator [Candes and Tao, 2007], we drop the constraint $\Theta \succ 0$ and consider the following optimization problem:

$$\begin{aligned} &\min \|\Theta\|_1 \quad s.t. \\ &\|(\Theta + \frac{\lambda_u}{\alpha} I)^{-1} - \frac{1}{d + \beta} (UU^\top + \beta\Sigma)\|_\infty \leq \frac{\gamma}{d + \beta}. \end{aligned} \quad (19)$$

By multiplying $\Theta + \frac{\lambda_u}{\alpha} I$ with the constraint, we obtain the following relaxation of Eq. (19):

$$\min \|\Theta\|_1 \quad s.t. \quad \|C\Theta - E\|_\infty \leq \tau, \quad (20)$$

where $C = \frac{1}{d + \beta} (UU^\top + \beta\Sigma)$, $E = I - \frac{\lambda_u}{\alpha} C$, and $\tau = \frac{\gamma}{d + \beta}$. This relaxation has been used in the constrained ℓ_1 -minimization for inverse matrix estimation (CLIME) [Cai *et al.*, 2011]. Indeed, the optimization in Eq. (20) is equivalent to the following optimization problem:

$$\min_{\Theta} \text{tr}(\Theta^\top C\Theta) - \text{tr}(E\Theta) + \tau \|\Theta\|_1. \quad (21)$$

Let $\tilde{\Theta}$ be the solution of Eq. (21), which is not necessarily symmetric. We use the following symmetrization step to obtain the final $\hat{\Theta}$,

$$\hat{\Theta}_{ik} = \hat{\Theta}_{ki} = \tilde{\Theta}_{ik} \mathbb{I}(|\tilde{\Theta}_{ik}| \leq |\tilde{\Theta}_{ki}|) + \tilde{\Theta}_{ki} \mathbb{I}(|\tilde{\Theta}_{ik}| > |\tilde{\Theta}_{ki}|), \quad (22)$$

where $\mathbb{I}(\cdot)$ is an indicator function, which is equal to 1 if the condition is satisfied, and otherwise 0. In addition, although constraint $\Theta \succ 0$ is not imposed in Eq. (21), the symmetrized $\hat{\Theta}$ is positive definite with high probability and converge to the solution of Eq. (17) under the spectral norm, guaranteed by Remark 1 and Theorem 1 in [Liu and Luo, 2014].

The optimization problem in Eq. (21) can be solved using the ADMM algorithm [Boyd *et al.*, 2011]. The augmented Lagrangian of Eq. (21) is as follows:

$$\begin{aligned} L_\rho(\Theta, Z, Y) = &\text{tr}(Z^\top C Z) - \text{tr}(E Z) + \tau \|\Theta\|_1 + \langle Y, \Theta - Z \rangle \\ &+ \frac{\rho}{2} \|\Theta - Z\|_F^2, \end{aligned} \quad (23)$$

Algorithm 1: PRMF Optimization Algorithm

Input : $\mathcal{D}, \Sigma, d, \lambda_u, \lambda_v, \alpha, \beta, \gamma, \theta, \rho$
Output: U, V, Θ

- 1 **if** $\beta > 0$ **then**
- 2 $X \leftarrow \text{SVD}(\Sigma, d)$;
- 3 Initialize U and V randomly, and set $\Theta = I$;
- 4 **for** $iter = 1, 2, \dots, \max_iter$ **do**
- 5 **for** $t = 1, 2, \dots, T$ **do**
- 6 **foreach** $(u_i, v_j, R_{ij}) \in \mathcal{D}$ **do**
- 7 $U_i \leftarrow U_i + \theta(\Delta_{ij}V_j - \lambda_u U_i - \alpha\Theta_{i*}U)$;
- 8 $V_j \leftarrow V_j + \theta(\Delta_{ij}U_i - \lambda_v V_j)$;
- 9 **if** $\beta = 0$ **then**
- 10 $\hat{U} = \frac{1}{\sqrt{d}} U$; $\tau = \frac{\gamma}{d}$;
- 11 **else**
- 12 $\hat{U} = [\frac{1}{\sqrt{d+\beta}} U, \frac{\sqrt{\beta}}{\sqrt{d+\beta}} X]$; $\tau = \frac{\gamma}{d+\beta}$;
- 13 $Z^0 = \Theta$; $Y^0 = 0$; $E = I - \frac{\lambda_u}{\alpha} \hat{U} \hat{U}^\top$;
- 14 $P = I - \frac{1}{\rho} \hat{U} (I + \frac{1}{\rho} \hat{U}^\top \hat{U})^{-1} \hat{U}^\top$;
- 15 **for** $t = 0, 1, \dots, K - 1$ **do**
- 16 $\Theta^{t+1} = \text{soft}[Z^t - Y^t, \frac{\tau}{\rho}]$;
- 17 $Z^{t+1} = P(\frac{1}{\rho} E + \Theta^{t+1} + Y^t)$;
- 18 $Y^{t+1} = Y^t + \Theta^{t+1} - Z^{t+1}$;
- 19 $\Theta = \Theta^K$;

where Y is a scaled dual variable and $\rho > 0$. We can obtain the following ADMM updates:

$$\Theta^{t+1} = \arg \min \frac{\rho}{2} \|\Theta - Z^t + Y^t\|_F^2 + \tau \|\Theta\|_1, \quad (24)$$

$$Z^{t+1} = \arg \min \frac{\rho}{2} \|\Theta^{t+1} - Z + Y^t\|_F^2 + \text{tr}(Z^\top C Z) - \text{tr}(E Z), \quad (25)$$

$$Y^{t+1} = Y^t + \Theta^{t+1} - Z^{t+1}. \quad (26)$$

The closed solution to Eq. (24) is as follows:

$$\Theta^{t+1} = \text{soft}[Z - Y, \frac{\tau}{\rho}], \quad (27)$$

where

$$\text{soft}[A, \lambda] = \begin{cases} A_{ij} - \lambda & \text{if } A_{ij} > \lambda, \\ A_{ij} + \lambda & \text{if } A_{ij} < -\lambda, \\ 0 & \text{otherwise.} \end{cases} \quad (28)$$

The solution to Eq. (25) is as follows:

$$Z = (\frac{1}{\rho} C + I)^{-1} (\frac{1}{\rho} E + \Theta + Y). \quad (29)$$

The time complexity of the inverse operation in Eq. (29) is $\mathcal{O}(m^3)$. As Σ is symmetric, to improve the computation efficiency, we consider the following approximation $\Sigma \approx X X^\top$, where $X \in \mathbb{R}^{m \times d}$. In this paper, we obtain X by factorizing Σ using SVD and choosing the singular vectors with respect to the top d largest singular values (see line 2 in Algorithm 1).

Let $\hat{U} = [\frac{1}{\sqrt{d+\beta}}U, \frac{\sqrt{\beta}}{\sqrt{d+\beta}}X] \in \mathbb{R}^{m \times 2d}$. When the prior covariance matrix Σ is not available, we set $\hat{U} = \frac{1}{\sqrt{d}}U$. Then, $C \approx \hat{U}\hat{U}^\top$. Following Woodbury matrix identity, we have

$$(\frac{1}{\rho}C + I)^{-1} \approx I - \frac{1}{\rho}\hat{U}(I + \frac{1}{\rho}\hat{U}^\top\hat{U})^{-1}\hat{U}^\top. \quad (30)$$

Using this matrix operation, the time complexity of the update of Z in Eq. (29) becomes $\mathcal{O}(dm^2)$. The details of the proposed optimization algorithm are summarized in Algorithm 1. At each iteration, the time complexity of the SGD updates is $\mathcal{O}(T \cdot |\mathcal{D}| \cdot \bar{m} \cdot d)$, where \bar{m} denotes the average number of nonzero elements in each row of Θ . Thus, the sparsity of Θ can help improve the computation efficiency of the proposed method. The time complexity of the ADMM updates is $\mathcal{O}(K \cdot d \cdot m^2)$. Empirically, we set $T = 30$ and $K = 30$ in our experiments.

4 Experiments

In this section, we conduct empirical experiments on real datasets to demonstrate the effectiveness of the proposed recommendation method.

4.1 Experimental Setting

Dataset Description

The experiments are performed on four public datasets: MovieLens-100K², MovieLens-1M³, Ciao, and Epinions⁴. MovieLens-100K contains 100,000 ratings given by 943 users to 1,682 movies. MovieLens-1M consists of 1,000,209 ratings given by 6,040 users to 3,706 movies. We denote these two datasets by ML-100K and ML-1M, respectively. For the Ciao and Epinions datasets, we remove the items that have less than 3 ratings. Finally, on Ciao dataset, we have 185,042 ratings given by 7,340 users to 21,881 items. On Epinions dataset, there are 642,104 ratings given by 22,112 users to 59,104 items. The densities of the rating matrices on these datasets are 6.30% for ML-100K, 4.47% for ML-1M, 0.12% for Ciao, and 0.05% for Epinions. Moreover, on Ciao and Epinions datasets, we have observed 111,527 and 353,419 social relationships between users. The densities of the social relation matrices are 0.21% for Ciao and 0.07% for Epinions. Table 1 summarizes the details of the experimental datasets.

For each dataset, 80% of the observed ratings are randomly sampled for training, and the remaining 20% observed ratings are used for testing.

Evaluation Metrics

The performances of the recommendation algorithms are evaluated by two most popular metrics: mean absolute error (MAE) and root mean square error (RMSE). The definitions

Table 1: The statistics of the experimental datasets.

	ML-100K	ML-1M	Ciao	Epinions
# Users	943	6,040	7,340	22,112
# Items	1,682	3,706	21,881	59,104
# Ratings	100,000	1,000,209	185,042	642,104
Rating Density	6.30%	4.47%	0.12%	0.05%
# Social Rel.	N.A.	N.A.	111,527	353,419
Social density	N.A.	N.A.	0.21%	0.07%

Table 2: Performance comparisons on datasets without users' explicit social relationships.

Dataset	Method	RMSE	MAE
ML-100K	PMF	0.9251±0.0021	0.7314±0.0017
	SR ^{imp}	0.9205±0.0013	0.7290±0.0008
	PRMF	0.9157±0.0006	0.7226±0.0005
	PRMF ^{imp}	0.9132±0.0003	0.7210±0.0005
ML-1M	PMF	0.8728±0.0029	0.6807±0.0023
	SR ^{imp}	0.8621±0.0008	0.6739±0.0006
	PRMF	0.8592±0.0009	0.6738±0.0010
	PRMF ^{imp}	0.8572±0.0009	0.6727±0.0010

of MAE and RMSE are as follows:

$$MAE = \frac{1}{|\mathcal{D}_{test}|} \sum_{(u_i, v_j) \in \mathcal{D}_{test}} |R_{ij} - \hat{R}_{ij}|, \quad (31)$$

$$RMSE = \sqrt{\frac{1}{|\mathcal{D}_{test}|} \sum_{(u_i, v_j) \in \mathcal{D}_{test}} (R_{ij} - \hat{R}_{ij})^2}, \quad (32)$$

where R_{ij} denotes the observed rating in the testing data, \hat{R}_{ij} is the predicted rating, and $|\mathcal{D}_{test}|$ is the number of tested ratings. From the definitions, we can notice that lower MAE and RMSE values indicate better recommendation accuracy.

Evaluated Recommendation Methods

We compare the following recommendation methods: (1) **PMF**: This is the probabilistic matrix factorization model introduced in Section 2.1; (2) **SR^{imp}**: This is the SR method that exploits users' implicit social relationships for recommendation [Ma, 2013]; (3) **SR^{exp}**: This is the SR method that exploits users' explicit social relationships for recommendation [Ma *et al.*, 2011]; (4) **LOCABAL**: This is the social recommendation model proposed in [Tang *et al.*, 2013], which exploits both local and global social context for recommendation; (5) **eSMF**: This is the extended social matrix factorization model that extends LOCABAL to consider the graph structure of social neighbors for recommendation [Hu *et al.*, 2015]; (6) **PRMF**: This method learns users' social dependency without prior information about users' social relationships. The objective function is Eq. (10); (7) **PRMF^{imp}**: This is the proposed method that exploits users' implicit social relationships as the prior knowledge used to learn users' social dependency; (8) **PRMF^{exp}**: This is the proposed method that exploits users' explicit social relationships to learn the social dependency between users.

Parameter Settings

We adopt cross-validation to choose the parameters for the evaluated algorithms. The validation data is constructed by

²<http://grouplens.org/datasets/movielens/100k/>

³<http://grouplens.org/datasets/movielens/1m/>

⁴<http://www.public.asu.edu/~jtang20/datasetcode/truststudy.htm>

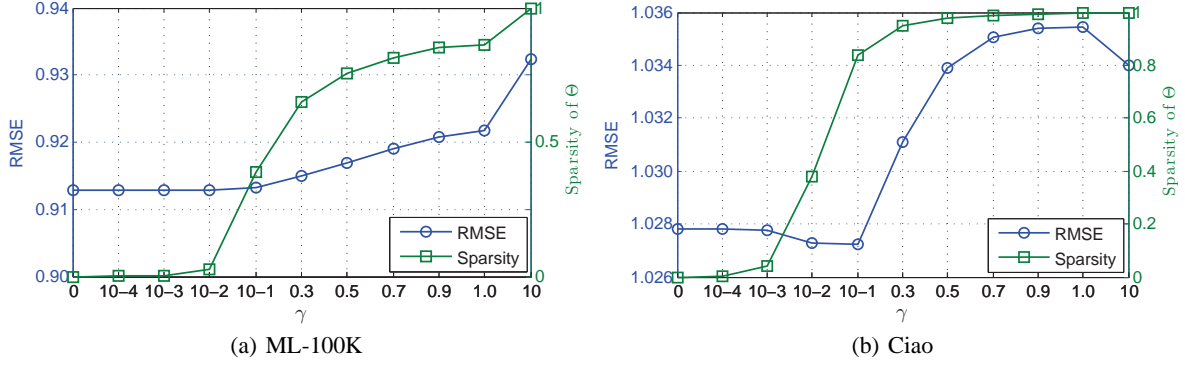


Figure 1: Performance trend of PRMF^{imp} on the ML-100K and Ciao datasets measured by RMSE with different settings of γ .

Table 3: Performance comparisons on datasets with users' explicit social relationships.

Dataset	Method	RMSE	MAE
Ciao	PMF	1.1031 ± 0.0045	0.8439 ± 0.0034
	SR^{imp}	1.0597 ± 0.0042	0.8234 ± 0.0036
	SR^{exp}	1.0286 ± 0.0029	0.7951 ± 0.0020
	LOCABAL	1.0777 ± 0.0039	0.8330 ± 0.0022
	eSMF	1.0592 ± 0.0043	0.8122 ± 0.0010
	PRMF	1.0326 ± 0.0030	0.8006 ± 0.0023
	PRMF^{imp}	1.0305 ± 0.0030	0.7993 ± 0.0023
	PRMF^{exp}	1.0258 ± 0.0035	0.7980 ± 0.0025
Epinions	PMF	1.1613 ± 0.0022	0.8932 ± 0.0019
	SR^{imp}	1.1321 ± 0.0054	0.8874 ± 0.0016
	SR^{exp}	1.1263 ± 0.0027	0.8795 ± 0.0023
	LOCABAL	1.1289 ± 0.0008	0.8734 ± 0.0010
	eSMF	1.1306 ± 0.0025	0.8749 ± 0.0025
	PRMF	1.1097 ± 0.0024	0.8703 ± 0.0022
	PRMF^{imp}	1.1081 ± 0.0023	0.8691 ± 0.0022
	PRMF^{exp}	1.1085 ± 0.0024	0.8695 ± 0.0022

randomly chosen 10% of the ratings in the training data. For matrix factorization methods, we set the dimensionality of the latent space d to 10. The latent features of users and items are randomly initialized by a Gaussian distribution with mean 0 and standard deviation $1/\sqrt{d}$. Moreover, we set the regularization parameters $\lambda_u = \lambda_v$ and choose the parameters from $\{10^{-5}, 10^{-4}, \dots, 10^{-1}\}$. For PRMF, α is chosen from $\{2^{-5}, 2^{-4}, \dots, 2^{-1}\}$, θ is chosen from $\{2^{-5}, 2^{-4}, \dots, 2^{-1}\}$. Moreover, we set $\gamma = 10^{-4}$, $\beta = 10$, and $\rho = 100$. For SR methods, the regularization parameter α is chosen from $\{2^{-7}, 2^{-6}, \dots, 2^0\}$. The user similarity is computed using Pearson correlation coefficient, and we set the threshold of the user similarity at 0.75 and $N = 10$, following [Ma, 2013]. The parameters of LOCABAL and eSMF are set following [Tang *et al.*, 2013] and [Hu *et al.*, 2015].

4.2 Summary of Experiments

Table 2 summarizes the experimental results on the datasets without users explicit social relationships, and Table 3 summarizes the results on the datasets with users' explicit social relationships. We make the following observations:

- On all datasets, PRMF outperforms PMF by 0.94% on ML-100K, 1.36% on ML-1M, 7.05% on Ciao, and 5.16%

on Epinions, in terms of RMSE. This indicates the recommendation accuracy can be improved by jointly learning users' preferences and users' social dependency.

- Compared with SR^{imp} , PRMF^{imp} achieves better results on all datasets. For example, in terms of RMSE, PRMF^{imp} outperforms SR^{imp} by 0.73%, 0.49%, 2.92%, and 2.40%, respectively. This shows that users' social dependency learned from the rating data is more effective than the user dependency predefined based on users' implicit social relationships, for improving the recommendation accuracy.
- The proposed PRMF^{exp} method outperforms the state-of-the-art social recommendation methods that exploits users' explicit social relationships for recommendation. On the Ciao and Epinions datasets, the average improvements of PRMF^{exp} over SR^{exp} , LOCABAL, and eSMF, in terms of RMSE, are 1.03%, 3.60%, 2.77%, respectively. This again demonstrates the effectiveness of the proposed method.
- PRMF^{imp} and PRMF^{exp} outperforms PRMF on all datasets. This indicates the prior information (e.g., users' explicit and implicit social relationships) are beneficial in improving recommendation accuracy. However, the improvements are not very significant. One potential reason is the SVD factorization used in Algorithm 1 (line 2) may not accurately approximate the prior covariance matrix.

In addition, we also study the impact of the sparsity of the learned social dependency matrix Θ on the recommendation accuracy. We choose the regularization parameter γ from $\{0, 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 10\}$. Figure 1 shows the performance trend of PRMF^{imp} on the ML-100K and Ciao datasets, in terms of RMSE. Observed that the sparsity of Θ increases with the increase of γ . As shown in Figure 1(a), denser social dependency matrix generally achieves better recommendation accuracy. However, denser social dependency matrix cause more computation time used to learn the user latent features. Indeed, there exists some balance between the computation efficiency and the recommendation accuracy. For example, on the ML-100K dataset, by setting γ to 0.3, the sparsity of the learned Θ is 65.21%, and the RMSE value is 0.9149, which is 0.56% better than the best competitor SR^{imp} . Moreover, Figure 1(b) also indicates that better recommendation accuracy may be achieved by learning a sparse Θ . For example, on the Ciao dataset, the best recommendation accuracy is achieved by set-

ting γ to 0.1, and the sparsity of the learned Θ is 79.74%. On the ML-1M and Epinions datasets, we have similar observations with that on the ML-100K dataset. Due to space limitation, we do not report those results here.

5 Conclusion and Future Work

In this paper, we propose a novel social recommendation method, named probabilistic relational matrix factorization (PRMF). For a specific recommendation task, the proposed approach jointly learns users' preferences and the optimal social dependency between users, to improve the recommendation accuracy. Empirical results on real datasets demonstrate the effectiveness of PRMF, in comparison with start-of-the-art social recommendation algorithms.

The future work will focus on the following potential directions. First, we would like to develop more efficient optimization algorithms for PRMF, based on the parallel optimization framework proposed in [Wang *et al.*, 2013]. Second, we are also interested in extending PRMF to solve the top- N item recommendation problems.

References

- [Boyd *et al.*, 2011] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.
- [Cai *et al.*, 2011] Tony Cai, Weidong Liu, and Xi Luo. A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.
- [Candes and Tao, 2007] Emmanuel Candes and Terence Tao. The dantzig selector: statistical estimation when p is much larger than n . *The Annals of Statistics*, pages 2313–2351, 2007.
- [Dhillon, 2007] JVDI Dhillon. Differential entropic clustering of multivariate gaussians. In *NIPS'07*, 2007.
- [Guo *et al.*, 2015] Guibing Guo, Jie Zhang, and Neil Yorke-Smith. Trustsvd: Collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In *AAAI'15*, 2015.
- [Hu *et al.*, 2015] Guang-Neng Hu, Xin-Yu Dai, Yunya Song, Shu-Jian Huang, and Jia-Jun Chen. A synthetic approach for recommendation: combining ratings, social relations, and reviews. In *IJCAI'15*, pages 1756–1762. AAAI Press, 2015.
- [Jamali and Ester, 2009] Mohsen Jamali and Martin Ester. Trustwalker: a random walk model for combining trust-based and item-based recommendation. In *KDD'09*, pages 397–406. ACM, 2009.
- [Jamali and Ester, 2010] Mohsen Jamali and Martin Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys'10*, pages 135–142. ACM, 2010.
- [Koren *et al.*, 2009] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.
- [Koren, 2008] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *KDD'08*, pages 426–434. ACM, 2008.
- [Li *et al.*, 2015] Hui Li, Dingming Wu, Wenbin Tang, and Nikos Mamoulis. Overlapping community regularization for rating prediction in social recommender systems. In *RecSys'15*, pages 27–34. ACM, 2015.
- [Linden *et al.*, 2003] Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *Internet Computing, IEEE*, 7(1):76–80, 2003.
- [Liu and Luo, 2014] Weidong Liu and Xi Luo. High-dimensional sparse precision matrix estimation via sparse column inverse operator. *Journal of Multivariate Analysis*, 2014.
- [Ma *et al.*, 2011] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *WSDM'11*, pages 287–296. ACM, 2011.
- [Ma, 2013] Hao Ma. An experimental study on implicit social recommendation. In *SIGIR'13*, pages 73–82. ACM, 2013.
- [Ma, 2014] Hao Ma. On measuring social friend interest similarities in recommender systems. In *SIGIR'14*, pages 465–474. ACM, 2014.
- [Mnih and Salakhutdinov, 2007] Andriy Mnih and Ruslan Salakhutdinov. Probabilistic matrix factorization. In *NIPS'07*, pages 1257–1264, 2007.
- [Su and Khoshgoftaar, 2009] Xiaoyuan Su and Taghi M Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009:4, 2009.
- [Tang *et al.*, 2013] Jiliang Tang, Xia Hu, Huiji Gao, and Huan Liu. Exploiting local and global social context for recommendation. In *IJCAI'13*, pages 2712–2718. AAAI Press, 2013.
- [Wang *et al.*, 2013] Huahua Wang, Arindam Banerjee, Choji Hsieh, Pradeep K Ravikumar, and Inderjit S Dhillon. Large scale distributed sparse precision estimation. In *NIPS'13*, pages 584–592, 2013.
- [Yang *et al.*, 2012] Xiwang Yang, Harald Steck, and Yong Liu. Circle-based recommendation in online social networks. In *KDD'12*, pages 1267–1275. ACM, 2012.
- [Yang *et al.*, 2014] Xiwang Yang, Yang Guo, Yong Liu, and Harald Steck. A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1–10, 2014.
- [Yu *et al.*, 2011] Le Yu, Rong Pan, and Zhangfeng Li. Adaptive social similarities for recommender systems. In *RecSys'11*, pages 257–260. ACM, 2011.

- [Zhang and Yeung, 2010] Yu Zhang and Dit-Yan Yeung. A convex formulation for learning task relationships in multi-task learning. In *UAI'10*, pages 733–742, 2010.
- [Zhang *et al.*, 2013] Xingang Zhang, Qijie Gao, Christopher S.G. Khoo, and Amos Wu. Categories of friends on social networking sites: An exploratory study. In *ALIEP'13*, 2013.